A Network Representation of Raster Land-Cover Patches

Lee De Cola

Abstract

Network models, based on mathematical graph theory, are used in many fields, from chemistry and biology to ecology and sociology. A nominal-valued land-cover raster can be represented as a network in which patches are nodes connected by links if two patches are adjacent. Built with existing GIS technology, the system is applied to the largest patches of a raster for the Washington, D.C. area in order to provide a data structure that visualizes the gross spatial structure of the raster; provides information about the overall arrangement of the patches and their adjacencies; and generates new metrics about land-cover patterns, connectivity and spatial autocorrelation. The system facilitates the evaluation of land-cover classifications, the measurement of landscape change, comparisons of data from different regions, and the analysis of land/environmental interactions.

Introduction

Land-Cover mapping

A land-cover (LC) raster categorizes the pixels of remotely sensed data (Richards and Jia, 2006) into a small set of nominal values (Wickham *et al.*, 1996) meant to characterize how the land appears or is being used. Pixel frequencies (counts) of these classes generally correspond to the composition (in the sense of Boots (2003)) of the land-covers of the ecosystems within which the data fall, and fieldwork suggests that the classifications are often quite accurate. These results demonstrate that LC technology provides representations of the underlying phenomenon that are useful in such fields as ecoregion characterization (Gallant *et al.*, 2004), change detection (Loveland *et al.* 2002), and resource management (Jones *et al.*, 2001).

Nevertheless, the analysis of raster data offers at least four daunting challenges. First, even simple systems can present many patterns. For example, the 3-valued 4×4 cell raster shown in the first frame of Figure 1 is but one realization from over 40 million possible configurations. A raster of size and dynamic range sufficient to capture real-world complexity (say 8-bit values measured over 1000^2 cells) will have a virtually infinite number of possible patterns. A second problem is "noise" in which the patch size distribution will usually be dominated by the smallest, especially single-cell, patches (Vogt *et al.*, 2007). Third, there are a large number of landscape characterization metrics applicable at multiple levels: although LC class characterizes the individual pixel, other measures proliferate as attention scales up through the neighborhood and patch to the LC class and ultimately the landscape level (Griffith *et al.* 2003). Finally, these and other statistics will naturally be highly sensitive to the chosen classification scheme, spectral specification, and resolution (Quattrochi and Goodchild, 1996).

The nearly boundless complexity of such rasters, coupled with their evident utility, has stimulated much research devoted to developing succinct summarizations of LC data, and we find that metrics can be arrayed among major dimensions of landscape characterization (see (Cushman *et al.*, 2008) for one such system). No single measure is ever sufficient because the data describe such landscape features as diversity, connectivity, fragmentation, and autocorrelation, to mention just a few characteristics. This paper uses the unique insights of network theory to present a way to transform an LC raster into a simple data structure that is easily visualized, retains much of the important thematic and spatial information of the original data, and can help LC analysts make rapid distinctions among related classes.

The paper is organized as follows. After an introduction to network theory and a brief review of its application to GIScience, the methodology and network characterization technique is presented. I then demonstrate how a 123 km^2 LC raster centered on the Northern Virginia region of the United States can be represented by a network which is then visualized, summarized in plots and tables, and characterized by a few statistics. The system is then used to compare two LC rasters from different dates, and a final section elaborates on applications, acknowledges some problems, and offers suggestions for future work.

Network Theory

In mathematics a *network* is a set of abstract elements called *nodes* that can be pairwise connected by *links*. (The mathematical terms for these entities are graph, vertex, and edge, respectively (Chartrand and Zhang, 2005).) For readers who may be unfamiliar with graph theory, a network representation of the United States is shown in the first frame of Figure 2, in which each region (51 states including the District of Columbia) is a node plotted at the longitude and latitude of the centroid, with a line segment linking two states if they share a common boundary. The descriptive statistics in the second frame report that the network has 51 nodes (vertices) and 108 links (edges), that two of the nodes are isolated (have no neighbors) and that 34 nodes border the exterior "region" (Canada, Mexico, or the ocean). This simple data structure can be analyzed to reveal a great deal

U.S. Geological Survey, 521 National Center, Reston, VA 20192 (ldecola@usgs.gov).

Photogrammetric Engineering & Remote Sensing
Vol. 76, No. 1, January 2010, pp. 61–72.

0099-1112/10/7601–61/$3.00/0

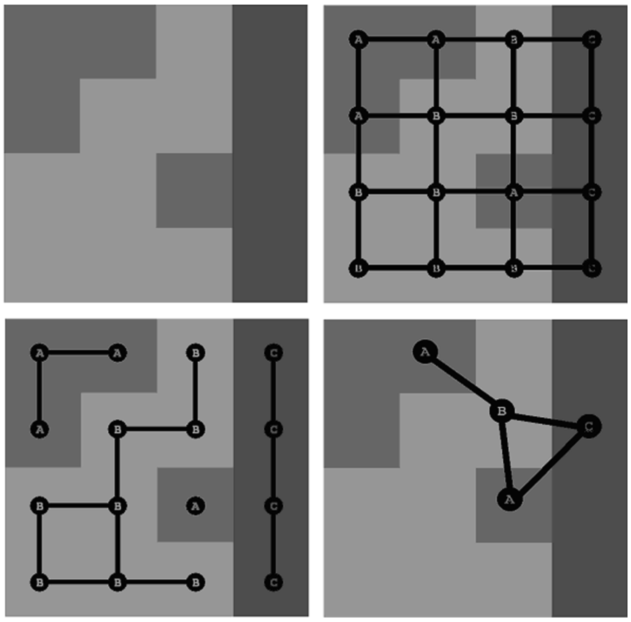


Figure 1. Basic operations to transform a three-color 4 × 4 raster into a simple network.

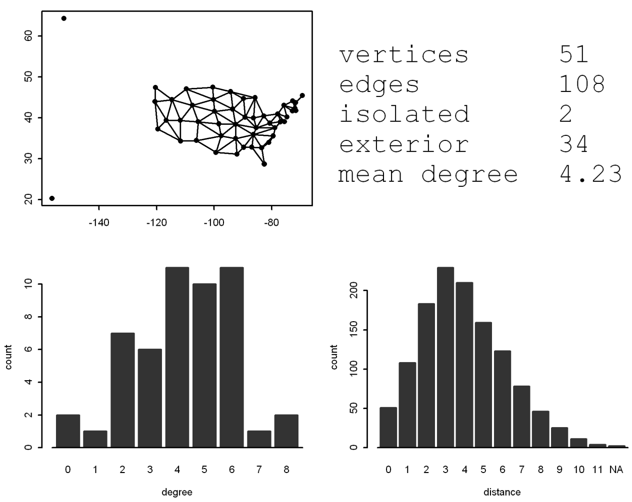


Figure 2. Report on the United States graph showing a plot of the network, its important statistics, and bar charts for degrees and distances.

about the state system. One important quality of a node is its *degree*, the number of links that connect it to adjacent nodes. For example, the isolated nodes Alaska and Hawaii shown to the west of the plot have no neighbors, so their degree = 0, while Missouri and Tennessee in the southeast each have the network's maximum degree = 8. In the third frame the bar plot of the frequencies of the degrees shows that most states have between four and six neighbors. An important quality of the network as a whole is the mean degree of the nodes, which in the case of the United States is $(2 \times 108)/51 = 4.23$ where the 2 accounts for the fact that every link connects two nodes.

It is possible to travel between any two of the 49 contiguous nodes of the US network along minimum length *paths* the number of whose links is the inter-node *distance*. This measurement, summarized in the bar plot of the fourth frame, provides intuitive metrics for the non-Euclidean structure of the network. For example, distance has a fairly symmetric distribution and a skewness = 0.476, and the individual bars in the plot summarize the network: at distance = 0 there are 51 nodes, at distance = 1 there are 108 links, distance = 3 is the mode, and there are 2 isolated nodes for which distance is undefined. Finally, the maximum "degree of separation" between any two nodes (e.g., the shortest path from California to Maine) is 11 links, and this longest inter-node path is appropriately called the *diameter* of the network. This four-frame exhibit format will be used below to summarize the LC network analyzed below.

A useful survey of network (graph) theory is Chartrand and Zhang (2005) and some of the powerful methods used here are implemented in Mathematica® (Pemmaraju and Skiena, 2003) and the R language (Gentleman, 2005). Although nodes and links may be portrayed (embedded in the plane) as points and lines, respectively, the network itself is an abstract mathematical entity whose arrangement is characterized by the above terms, which have precise definitions that, although intuitive, may not correspond exactly to the same or similar words in other fields.

Although the relevant fundamentals have recently been set forth in Worboys and Duckham (2004), GIScience has long recognized the applicability of networks to data structures. Relatively early practical applications are the original USGS digital line graph (DLG) vector data structure (Guptill, 1990) and Samet's (1989) development of raster quadtrees (Hinton and Salakhutdinov, 2006). Networks are also rapidly being adopted in environmental science to model such phenomena as organism interactions, food webs, habitat connectivity, migration, invasives, and so forth. For a useful review see Fortin and Dale (2005), and perhaps the most innovative example is the work of Bunn *et al.* (2000) and Urban and Keitt, (2001), who organize Mexican spotted owl habitat patches into a network to demonstrate the importance of connectivity among on the viability of a metapopulation. Finally, the present research furthers ongoing work in land-cover pattern characterization and especially patch morphology *(Riitters et al.*, 2006) size, shape, arrangement, diversity, and autocorrelation, (see Vogt *et al.*, (2007) for an example of forest patch analysis). In the language of current research (e.g., the influential FRAGSTATS system of Cushman *et al.*, (2008)), the present approach provides class-level metrics of neighborhood structure and fragmentation while suggesting new information about interactions among classes.

Methods

Figure 1 above shows how the system transforms a 4 × 4 raster each of whose cells may have one of three nominal (grayscale) values. In the second frame, each cell is linked to its orthogonal neighbor (diagonal links could also be examined, but results for large rasters would not be much influenced by this choice). In frame three, any joins that cross values are dissolved, so that four homogeneous patches are formed, and in the final frame each labeled patch is joined to each of its neighbors by a link in order to create the network "dual" of the patch map. In this "toy" example, 16 cells are transformed into a planar network of four nodes and four links. Although many configurations of this simple raster model would give the same network (consider what happens when the colors of a few cells are changed) differences in spatial structure among larger datasets will more strongly influence the kinds of networks generated.

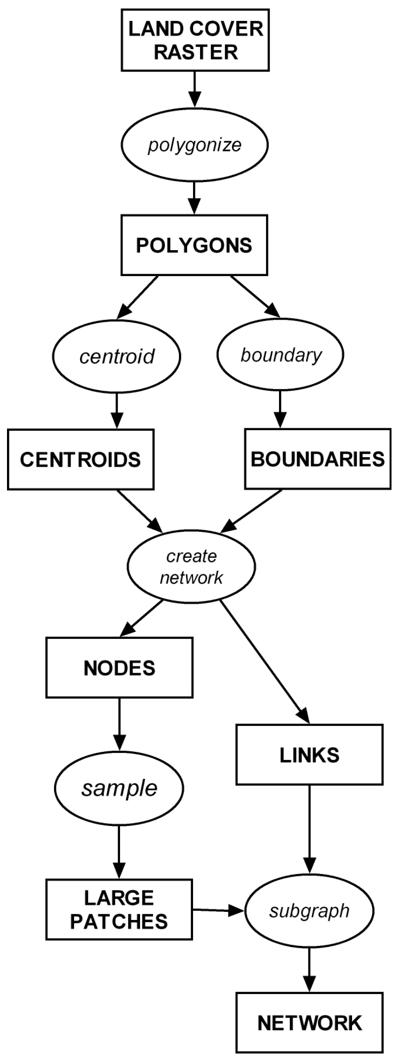


Figure 3. Simplified flowchart of the basic raster-to-network operations; data are in boxes and transformations in ovals.

Figure 3 outlines the transformation of a raster into a network. The top part of the figure represents the steps shown in Figure 1 in which patches become polygons whose centroids and boundaries are respectively the nodes and links of a network representing the spatial arrangement of *all* the patches in the raster. Because (as we shall see) typical rasters tend to have very many small and often even single-cell patches, the bottom part of the figure has a step that samples the largest nodes of the network based on the patch size distribution.

The system is built on standard GIS transformations (Turner *et al*., 1998), statistical calculations (de Smith *et al*., 2007), and network operations (Chartrand and Zhang, 2005). I use the ArcGIS® tools (Environmental Systems Research Institute, 2006) and S-PLUS® objects and functions (Insightful Corporation, 2001; Chapter 12) listed in Table 1, but other robust GIS and statistical software can carry out these operations. Moreover, the resulting network can be exported to other data formats: ArcGIS® for visualization and network analysis and Mathematica® for symbolic analysis (Pemmaraju and Skiena, 2003).

Ignoring the LC codes for the moment, a network is most simply described by the counts of its nodes and links, summarized in the second frame of Figure 2 for the United States. For a land-cover raster a small number of large patches suggests positive spatial autocorrelation as the values "clump" together rather than occur at random (Hilhorst, 2005); whereas many small patches suggests negative autocorrelation (De Cola, 1989) and (Boots, 1999). The number of links reflects the "connectedness" of the network: for a given number of nodes more links signifies more connectivity among patches. Beyond these two summaries, the size distribution of the degrees of the nodes (the third frame of Figure 2) tells us which nodes are more connected to or isolated from the others. Finally, the distribution of internode path distances, shown in the fourth frame, reflects the connectivity of the network. Such measures apply to any network, but the power of the method is revealed when we relate these descriptions to a LC raster, its individual patches, and its constituent classes, in order to provide insights into such problems as patch size scaling, sampling, and landscape fragmentation (Riitters *et al*., 2006) (Ferraz *et al*., 2007).

Results

The Full LC Network

I have chosen a raster from the Mid-Atlantic Land Cover Change study, which compiled data for the US states that intersect the Chesapeake Bay watershed (Jones *et al*.,

Table 1. A Few of the More Important ArcGIS® and S-PLUS® Functions used in the Analysis

GIS command (ARGUMENT)	EXPLANATION
RasterToPolygon (RASTER)	CONVERT RASTER TO POLYGON
CalculateField (TABLE)	COMPUTE AREA, PERIMETER, CENTROID COORDINATES
MakeXYEventLayer (TABLE)	TRANSFORM CENTROID COORDINATES TO POINTS
PolygonToLine (POLYGON)	CREATE BOUNDARIES
S-PLUSfunction (ARGUMENT)	**EXPLANATION**
CreateNetwork (TABLES)	CREATE A NETWORK OBJECT FROM GIS TABLES
Plot (NETWORK)	PLOT THE NODES AS POINTS AND LINKS AS LINES
Summary (NETWORK)	PRINT NETWORK ORDER, SIZE, ISOLATED NODES
Report (NETWORK)	OUTPUT TABLES DESCRIBING NETWORK STATISTICS
Adjacency (NETWORK)	COMPUTE THE ADJACENCY MATRIX
Distance (MATRIX)	COMPUTE THE DISTANCE MATRIX
InduceSubgraph (NETWORK)	SELECT NODES AND VALID LINKS
Export (NETWORK)	EXPORT LINKS AND NODES TO OTHER FORMATS

2001). The data represent seven land-cover values (introduced below) for 120-meter pixels aggregated from 1973 North American Land Cover (NALC) and 1992 National Land Cover Data (NLCD) (Environmental Protection Agency, 2002). From this dataset, a 1024^2 pixel block was extracted shown in Figure 4, a 123 km^2 study area that includes all of the Middle Potomac River watershed and the District of Columbia as well as parts of Northern Virginia, Maryland, and West Virginia.

Polygonization applied to this LC raster yields a network of 52,161 nodes (patches) and 84,494 links, for a mean degree = 3.24, i.e., on average a patch has about three neighbors. But these patches have a strongly positive skewed size distribution, as illustrated in the Pareto or "rank-size" plot shown in Figure 5 in which the areas of the patches are plotted against their ranks from largest (rank = 1) to smallest (single pixels), (Bettencourt *et al.*, 2007). The left-most point in the plot represents the 119,394 adjacent pixels of the largest (FOREST) patch in the study area, comprising 11.4 percent of the raster. This dominance is typical of systems in which one cell type "percolates" across the whole window (Stauffer and Aharony, 1992). At the other end of the size distribution are the 29,737 isolated pixels representing small (1.44 hectare) patches on the landscape as well perhaps as various kinds of "noise" in the data. Although these single-pixel "atomic" patches comprise only 2.84 percent of the area of the raster, they account for 57.0 percent of the nodes in this unsampled network.

In spite of a few kinks in this rank-size plot (almost certainly due to the truncation of the largest patches by the arbitrary study area window) the rank-size relationship is nearly linear, with a Pareto scaling exponent of −1.14 (Batty, 2005). In the network, a node represents each patch, from the largest to the smallest and ranging in area over five orders of magnitude. Although the raster itself is dominated by the *largest* contiguous patches, the resulting network is dominated by the *smallest* patches; and this extreme skewness is a challenge to land-cover analysis for three reasons. First, *inter alia*, we wish to give less attention to the smallest and usually less important patches in a raster. Second, the present report is intended to illustrate the network transformation with visualizable results applied to a reasonable number of nodes. Finally, from a relatively simple raster the system has generated a combinatorially complex network, so there is motivation to make the system more computationally manageable.

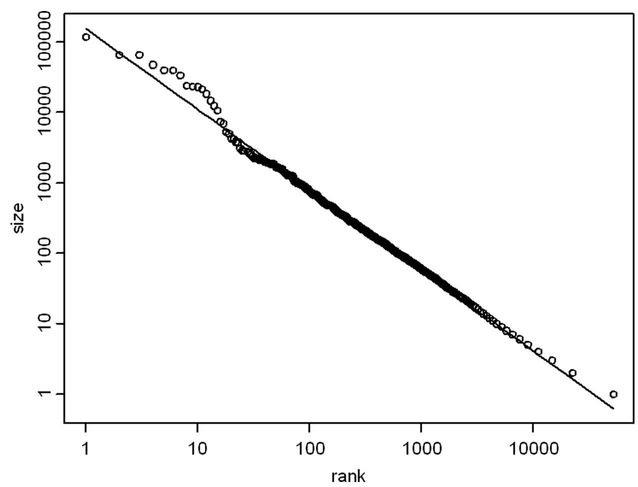


Figure 5. Rank-size plot of patch sizes for the full network on a log-log scale.

A method is therefore needed to filter the data in order to focus on the gross spatial structure of a network representing not only the largest patches but also the land-cover diversity of the raster (Wickham *et al.*, 2007). The system therefore samples from the full network (see Figure 3) as follows: (a) the patches are ranked as in Figure 5 from largest (rank = 1) to smallest (rank = 52,161), (b) mean area per patch is computed for each LC class, giving a weight that is higher for classes with relatively fewer patches, (c) this weight is used to reorder the rankings, raising the ranks of the less areally common LC classes, and then (d) the largest re-ranked patches (in the present case up to n = 1024) are selected to provide a stratified sample. The effect of this sampling is shown in Figure 6, in which the cumulative area of the raster is plotted against the new patch ranks. The first

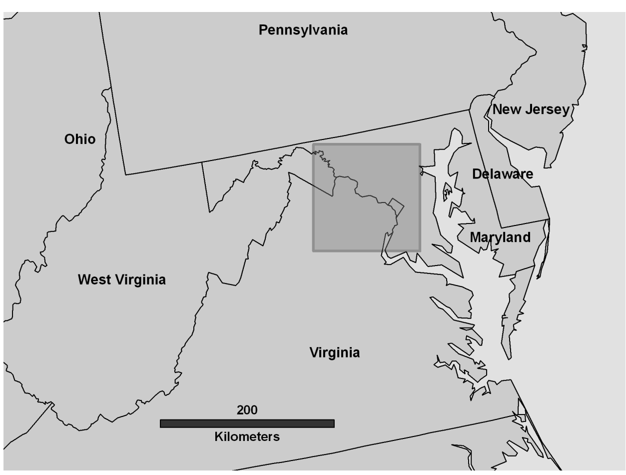


Figure 4. Study area within the Chesapeake Bay watershed including northern Virginia, central Maryland, and the District of Columbia

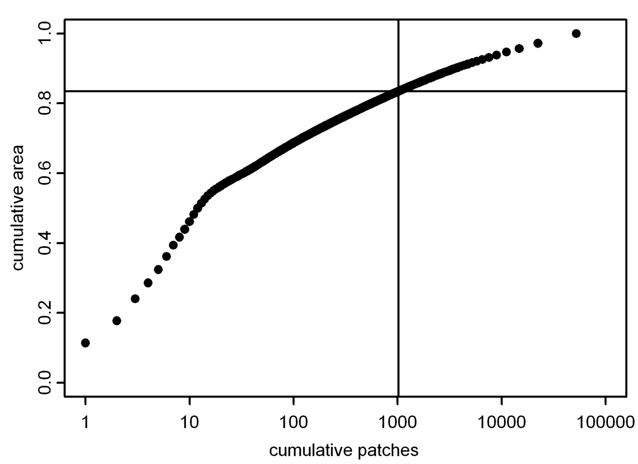


Figure 6. Cumulative patch areas as proportion of the raster plotted against the cumulative number of patches. Lines indicate that the largest 1,024 patches account for 83.4 percent of the raster.

point is the largest patch, comprising 11.4 percent of the raster; the second point is the next largest patch, contributing an additional 6.3 percent, and so forth. Note how much the largest patches contribute to the area of the raster as well as how little the smallest do: stratified sampling among the largest patches quickly accounts for a lot of area. Because only those nodes representing the largest patches are selected (in this case 1,024), any links that do not connect these nodes must also be deleted, a process known as subgraph induction (Chartrand and Zhang, 2005). For example, if the three Pacific coast states were removed from the US network, then the five links between them and their neighbors would also have to be deleted.

There is of course some arbitrariness to this sampling scheme. The removal of the smallest patches beyond a certain rank, technically called "censoring" (Helsel and Hirsch, 1992), will strongly bias against those LC classes with smaller patches. Stratification (the use of LC statistics to re-rank the patches) is designed to represent the diversity of the LC raster's land-cover classes, the nodes of its largest patches, and the links associated with boundaries between the sampled patches. This filtering is largely dictated by the immediate constraints of preliminary resources, but other methodologies could examine either the complete graph or other kinds of samples. Depending on computational resources, analytical purposes, or managerial needs, more and smaller patches would result from a higher rank cutoff up to the maximum, in which case weighting would be unnecessary (for another approach to patch sampling see Sohl *et al.* (2007)).

Sampled Network

Figure 7 describes the network composed of the stratified sample of 1,024 nodes (from the original 52,161) and the remaining 1626 links (from the original 84,494). As in Figure 2, the patches become points plotted at centroids and adjacencies become lines linking neighboring patches. Because the US state configuration is so well known, a comparison is illustrative. Not only does the LC network have many more nodes and links, it forms a much less regular spatial system than does the geopolitical network; yet, in spite of their differing sizes and structure, the mean links per node for both is about three. Nevertheless, the LC degree bar plot (frame 3) is very strongly skewed, with modal degree = 1 (roughly a third of the patches are "satellites" of a larger patch), and although most of the nodes have but a few neighbors, some have very many, up to maximum degree = 152. (Note that degree is shown not as a histogram but as a bar plot with irregularly spaced horizontal labels.) The links in effect are "oligopolized" by many fewer nodes, because degree is determined not only by the areas of the dominant patches but by important non-geometric factors, as we shall see. Furthermore, the shape of the distance histogram (shown in the fourth frame) is not so very different from that of the US network; both are fairly symmetrical and have similar maxima, and in fact this variable is even less skew (skewness = 0.180) than in the case of the US. Finally, the diameter of the LC network (the maximum distance between nodes) is 13 as compared to 11 for the US network. In summary, these exhibits are useful visualizations and reports on key metrics, indicating how the LC network reflects the spatial complexity of a landscape in contrast to the policy-imposed regularity of the US geopolitical system.

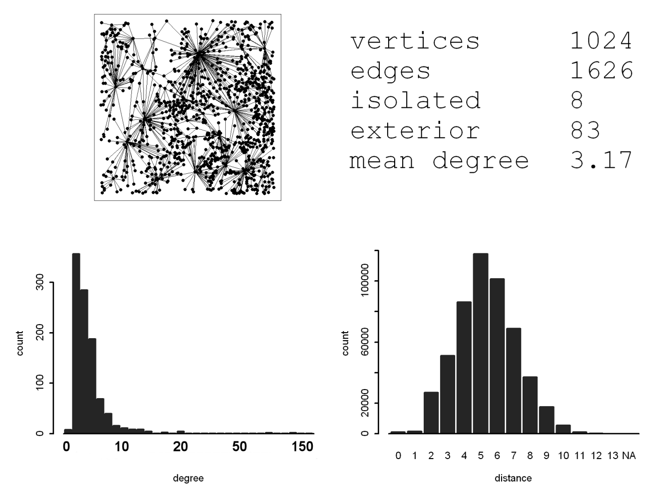


Figure 7. Report on the 1990 NLCD land-cover network showing a plot of the network, its important statistics, and bar plots for degrees and distances.

Although the spatial arrangement of LC values of the raster cells determines the structure of the network, we have not yet associated these categorical values with the network elements themselves. Plate 1 uses a GIS to display the source raster underneath both the nodes (symbolized as disks at patch centroids and colored according to the raster palette) and the associated links represented as an ArcGIS® network. Embedding the network in geographic space reveals important patterns. Note for example the centrality of the largest patches within "webs" created by their high degree (numbers of links). Particularly striking to the west are the large agriculture regions and those of the forested Blue Ridge, and, to the east, the Washington/Baltimore urban sprawl, as well as the Upper Potomac River estuary. In fact, the network structure associates the LC raster field with its salient constituent objects and their spatial relations, addressing a significant challenge in GIScience (Cova and Goodchild, 2002).

When the network nodes are partitioned by the seven LC classes, as in Table 2, we see important differences among the spatial attributes of the land-cover categories, of which four are most salient.

1. The bottom row of the table shows that although the 1,024 nodes represent just 1.96 percent of the 52,161 patches in the full network, they account for 82.6 percent of the pixels in the raster (recall Figure 6), i.e., the sampling is extremely efficient.
2. Comparing the areas and numbers of nodes of the land-covers (third and fourth columns of the table) shows how stratification provides a sample of nodes less skewed by class than by area; for example, although the last three classes together comprise less than 2 percent of the area, they account for 28.1 percent of the 1,024 the nodes. In spite of having removed over 92 percent of the patches, a richer collection of variously classed nodes remains.
3. The two areally largest classes AGRIC and FOREST each comprise about a third of the raster (column 3), but the FOREST class alone constitutes the plurality of the nodes (column 4) because its greater fragmentation results in its being more connected to other types of land-cover (hence as well its larger number of links). In the words of *Riitters et al.* (2002) this LC class constitutes the spatial (indeed the historical) "matrix" within which other covers arise.
4. Turning to the analysis of the network itself, the last two columns show that area and numbers of nodes do not completely determine numbers of links or degree: although WATER accounts for the smallest number of nodes, it has the highest mean degree of any of the LC classes because of its low fractal dimension and extreme "affinity" for some of the other classes; this behavior is explored below.

In sum, the LC raster has become an easily visualized data structure characterized by a few essential statistics.

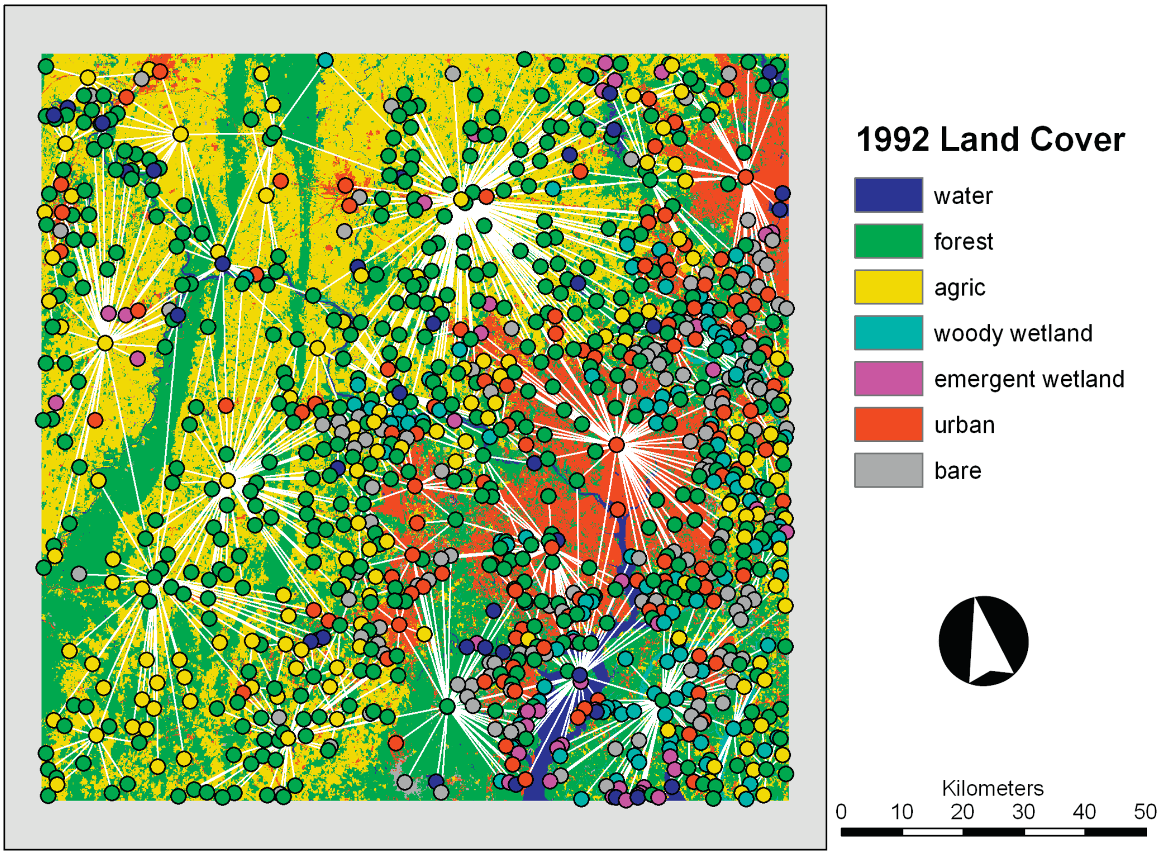


Plate 1. The LC network overlaid on the land-cover raster; nodes are symbolized with the same color scheme as the LC codes.

TABLE 2. DESCRIPTIVE STATISTICS OF LAND-COVER CLASSES FOR THE SAMPLED 1,024-NODE NETWORK

Land-Cover	Area	Percent of Total Area	Nodes	Links	Degree
AGRIC	365,495	34.9	132	718	5.44
FOREST	331,510	31.6	455	1212	2.66
URBAN	128,456	12.3	113	509	4.50
WATER	20,544	2.0	36	203	5.64
BARE	9,890	0.9	136	269	1.98
WOODWET	8,604	0.8	106	256	2.42
EMERGWET	798	0.1	46	85	1.85
TOTAL	**865,297**	**82.6%**	**1024**	**3252**	**3.17**

Autocorrelation on the Network

This network is next examined for its spatial autocorrelation, which describes the effect of distance on variation, a central concern of geographic analysis. Recall that the distance between two nodes is not the Euclidean measurement but the number of links in a minimum path between them. The fourth frame of Figure 7 illustrates that on the LC network this distance ranges from 0 (a node is zero links from itself!), though 1 for neighboring nodes, up to the network diameter = 13, with the 8 isolated nodes having an undefined distance within the network. The remainder of this section considers examines autocorrelation at distances = 1 and 2, and then the ensemble of all distances in the range [1, 13].

Because of the way the LC network was constructed, two nodes separated by distance = 1 are first-order neighbors, which raises the natural question "what is next to what?" Each of the 1,626 links represents an instance of a pairing of LC codes from among the $\binom{7}{2} = 21$ allowable LC pairs shown in the lower-triangular cells of the matrix in Table 3. It is clear that because the FOREST and AGRIC classes dominate the raster, the links that connect them will be the most frequent pairs. At the other end of the frequency spectrum, certain pairings (e.g., BARE/EMERGWET, BARE/WATER, and URBAN/EMERGWET) are rare, again in part reflecting the low frequencies of these numbers of these nodes.

These distance-1 LC pair frequencies are an integer-valued index of "nearness" in the sense that two land-covers that are frequently adjacent are "closer" than two kinds that

TABLE 3. COUNTS OF NUMBERS OF LINKS BETWEEN LC CLASSES FOR THE 1,626 LINKS

	FOREST	BARE	AGRIC	URBAN	WOOD-WET	EMERG-WET
BARE	135					
AGRIC	516	39				
URBAN	291	76	78			
WOODWET	130	16	41	42		
EMERGWET	30	1	12	5	11	
WATER	110	2	32	17	16	26

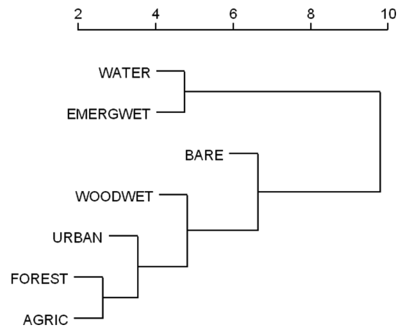


Figure 8. Tree in which height represents the order in which classes are grouped based on their adjacency frequencies.

are rarely so. These indices are used in hierarchical clustering to produce the dendrogram of Figure 8, where the scale represents the rescaled distance; e.g., WATER and EMERGWET tend to be about four nodes apart (De'ath and Fabricius, 2000). Reading the tree from left to right suggests the order in which LC classes could be merged: e.g., FOREST and AGRIC are most strongly associated; and indeed typically *are* difficult to separate in the classification process (Environmental Protection Agency, 2002). Alternatively, a right-to-left view of the tree highlights two major LC groupings: "wet" versus "dry" and within the dry group BARE standing apart from land covered by natural, cultivated, and cultural systems. This clustering not only summarizes the spatial association of the land-covers, but also visualizes how various alternative classification schemes might separate important classes.

Turning next to distance = 2, by the construction of the network, two nodes separated by this distance may have the *same* LC value, but they must be separated by nodes of a *different* value from each. Autocorrelation at this scale can therefore be summarized by the frequencies that nodes of one class are separated by nodes of other classes. In the LC patch context, we may say that one LC type "intervenes" between another type, and there are $7 \times 7 - 7 = 42$ such situations, reflected in the non-empty cells of Table 4. These distance-2 LC "joins" will be influenced not only by spatial "affinities" among the classes but also by the sizes of the associated patches (other things equal, large patches will tend to be next to other large patches). Because these counts are strongly heteroscedastic, I have regressed the logarithm of their frequencies on the logarithms of patch areas, and the table shows the residuals from this model in order to highlight which LC class tends to be found more or less often between nodes of a given class ($R^2 = 0.648$). For each LC in a table's row the number in the cell reflects how frequently (positive) or rarely (negative) that class intervenes between nodes of the LC in the column. For example, the first cell of the second row of the table represents the relative frequency of WATER-FOREST-WATER "triplets," which occur about exp(1.9) = 6.7 times as frequently as expected given the areas of the two LC classes. Further down the column, we see that WATER is most frequently fragmented by EMERGWET (as we might expect from Figure 8) but rarely by BARE (about one-tenth as frequently as would randomly be predicted). On the other hand URBAN patches are rarely separated by WATER, and examining the second row shows that FOREST often intervenes between AGRIC patches, and so forth. This relatively simple table summarizes much of the fragmentation of the original raster by reporting the extent to which certain land-covers "interrupt" others (Riitters *et al.*, 2006).

Classical autocorrelation analysis, in the sense of Moran and Geary (Haining, 2003), could have been used in the initial description of the raster, but is not appropriate here because node-creation has removed all same-class links (joins) within the patches. Nevertheless, as a statistical diagnostic the table highlights significant residuals, based on a *t*-test using their standard errors (Venables and Ripley, 1999). Note that this discussion relates to the well-known "four color problem" of determining the minimum number of nominal values that can be assigned to the nodes of a network so that adjacent nodes have different values (Chartrand and Zhang, 2005); in the state network, if Washington and California are both colored red then Oregon must have some other color. Although not developed here, this classic research theme will shed new light on the formal science of land-cover classification.

Finally, we may examine the co-occurrences of LC classes at *any* distance up to the network diameter = 13, and indeed this is the essence of spatial autocorrelation on a network. The fourth frame of Figure 7 is a bar chart of the $\binom{1024}{2} = 523{,}776$ internodal distances (Chartrand and Zhang, 2005), whose near symmetry is similar to that of the US inter-node distances of Figure 2. Moreover, as the LC adjacency frequencies were presented as a tree showing which classes were closest to which, so multidimensional scaling is again used, but now to map the inter-node distances onto a two-dimensional space to see which nodes are nearer to or further from others. In Figure 9 disks representing the 1,024 nodes are located at (x, y) coordinates representing the two principal eigenvalues

TABLE 4. INTERVENING CODES IN THE NETWORK. COLUMNS INDICATE LC CODES THAT ARE SEPARATED BY THE LC IN THE ROWS. SHOWN ARE RESIDUALS FROM A REGRESSION OF ADJACENCY FREQUENCIES ON NODE FREQUENCIES

	WATER	FOREST	AGRIC	WOODWET	EMERGWET	URBAN	BARE
WATER		−0.8	−1.2	−1.8*	0.8	−2.6**	−1.9*
FOREST	1.9*		2.9**	1.3	0.1	1.9*	0.1
AGRIC	−0.5	0.3		−0.1	−1.3	−1.5	−1.4
WOODWET	1.7	1.4	−0.8		1.2	0.2	−0.4
EMERGWET	3.1**	0.5	−0.4	−0.2		−1.7	−0.3
URBAN	0.4	0.1	−0.8	−0.6	−0.8		−0.1
BARE	−2.2*	1.3	−1.4	1.6	0.5	1.4	

*$p \le .10$, **$p \le .05$ (two-tailed test).

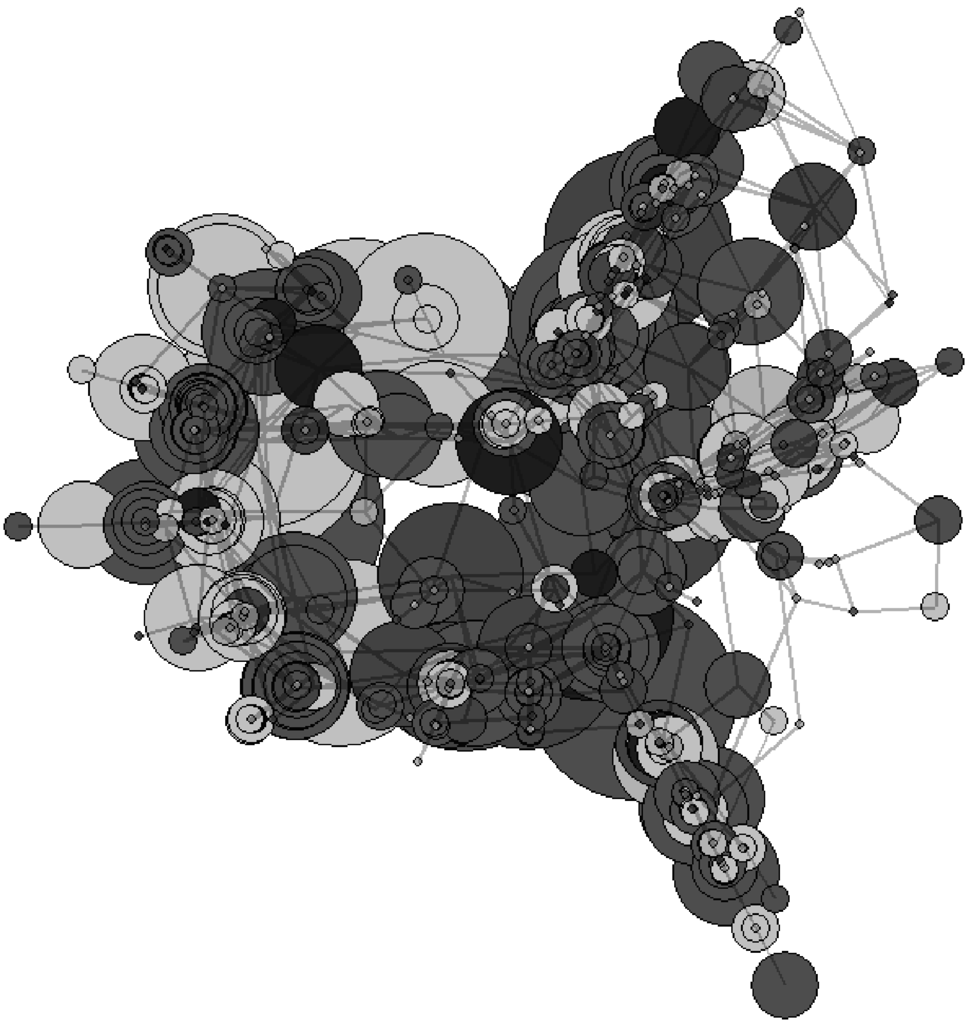


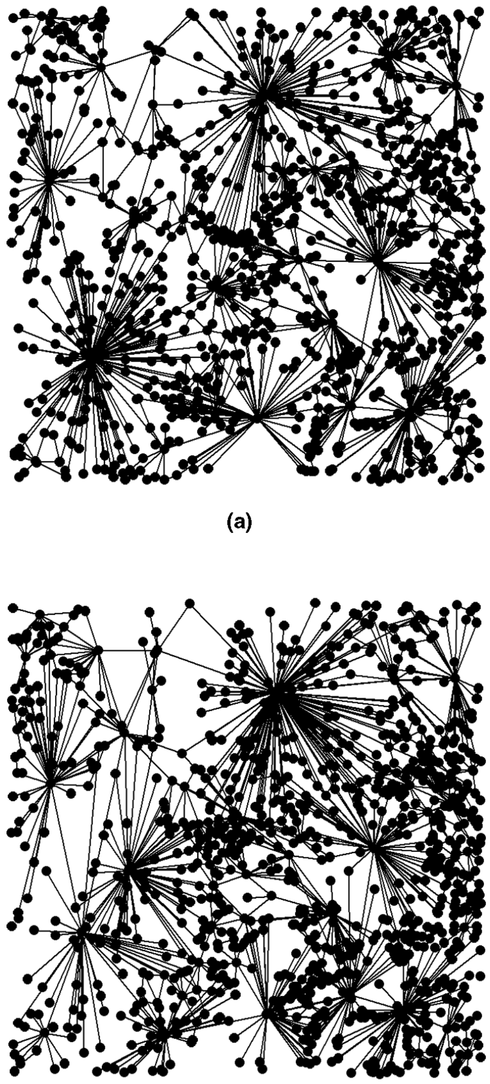
Figure 9. A visualization of the LC network in which patches are represented as disks whose size is proportional to area, color is that of the patch, and centers are at a point in the new space based on network distances.

of the rescaling ($r_{x,\mathrm{EAST}}$ = 0.921 and $r_{y,\mathrm{NORTH}}$ = 0.848), which together account for 61.4 percent of the total variance in the original Euclidean space. The disks are sized in proportion to the logarithms of the patch areas and symbolized on a grayscale corresponding to the LC codes to give a suggestion of the diversity of the network. This new "map," although quite unlike the original raster, reveals important patterns about the spatial structure of the abstract LC network (Pascual-Hortal and Saura, 2006). High-degree patches having many neighbors are generally central and those that are smaller and more isolated are peripheral, yet there remains a void occupied by no patch. Moreover, spatial variation in fragmentation is reflected in the way that "peninsulas" tend to form, most obviously in the estuarine and more urbanized east of the study area (see Plate 1). As a summary of the landscape, we see fragmentation due to hydrologic and urban processes within a visualization that incorporates four key elements of the underlying network: LC codes, the sizes of the patches, the "centrality" of the larger patches, and the general associations among LC classes. The raster field again gives forth new objects.

Change Analysis

Just as autoregression analyzes change among neighboring values on a linear temporal network (Hyndman *et al.*, 2002), so our spatial network can visualize and characterize LC change. In Figure 10 the analysis presented above for 1992 data is applied to an earlier 1973 scene (Environmental Protection Agency, 2002) for the study area of Figure 4. Although there are fundamental similarities between the networks, closer examination reveals important differences, such as a greater density of nodes to the east of the scene and the replacement of the high-degree (FOREST) node to the southwest with two lower-degree nodes.

These visual impressions are reinforced by the statistics of Table 5; the full networks are compared in the first three rows and the sampled networks in the last two rows. Note first that the 1992 dataset has 16.5 percent more nodes and 27.9 percent more links, so that the mean degree has risen almost 10 percent, suggesting how fragmentation results in smaller patches within a denser network. In keeping with this fragmentation is a considerable and statistically significant decline in the mean distances among the 1,024 sampled nodes of the two networks. These visual and statistical



(a)

(b)

Figure 10. Comparison of sampled networks for NALC and NLCD land-cover rasters: (a) 1973, and (b) 1992.

TABLE 5. BASIC STATISTICS COMPARING TWO LC NETWORKS FOR THE SAME STUDY AREA

	Raster Dataset	
	1973	1992
FULL NETWORK:		
Number of nodes	44,757	52,161
Number of links	66,055	84,494
Mean degree	2.95	3.24
SAMPLED NETWORK:		
Mean distance	5.58	5.34
Standard deviation	1.93	1.82

contrasts demonstrate further that the network representation is robust in its detection of important differences that are not overly sensitive to minor changes. The technique also suggests how comparisons between networks can also be applied to LC rasters from different locations.

Discussion

The application of network theory to land-cover raster analysis is quite natural. In the present case, a 1 megabyte raster has been reduced first to a network of 52,161 nodes and 84,494 links, and then to a sampled "skeleton" network of 1,024 nodes and 1,626 links. The nodes are simplifications of homogeneous patches, the links elegantly capture fundamental information about LC adjacency and spatial autocorrelation, and many of the tools from network theory (e.g., contraction, dualization, induction, coloring, connectivity) have natural applications to this practical field, leading to various kinds of insights:

1. Visualizations of the **network** (Figure 7 and Plate 1) capture important structural patterns among the largest patches in the raster, enabling ready qualitative comparisons across space and time (Figure 10).
2. Patches comprising 82.6 percent of the raster area are represented by 1,024 **nodes** whose distribution by LC code enables us to see which land-covers account for proportions of both LC areas and patch sizes (Table 2).
3. The 1,626 network **links** represent LC incidences whose degree counts show which kinds of land-covers are more connected than others (Table 2) and whose LC adjacencies provide a tree (Figure 8) that tells us which land-covers are more frequently neighbors.
4. Spatial autocorrelation on the network provides insights into fragmentation as a process in which patches are broken up by land-covers of other types (Table 4).
5. The network distance metric can be used to create a rescaled map of the nodes that shows the arrangement of patches in an adjacency space while retaining much of the original geographic structure (Figure 9).

Next Steps

Five next steps are proposed. First, although it is always exciting to apply new ideas to empirical data: only formal theory and controlled simulation can rigorously advance scientific understanding. The system is therefore being applied to simulated rasters with various degrees of spatial autocorrelation to test the power of the statistics as well as to estimate relationships between various network metrics and such parameters as raster size, value cardinality, and degrees of spatial autocorrelation (Turner *et al.*, 1998).

Second, an equally critical issue is the role of the classification system itself upon network structure. Land-cover classification maps each pixel onto a small nominal range, so there is interest in such questions as: how can the sampling process focus on larger patches and connected components?; what are the important characteristics of networks based on functional aggregations of classes (e.g., impervious covers (Yang *et al.*, 2003)) as well as new adjacency-based classes (e.g., riparian zones (Jones *et al.*, 2008))?; and how are these classifications reflected in such key measures of connectedness as order of the largest component, number of components, diameters, (Urban and Keitt, 2001)? Given the infinite ways that LC data can be created, information on the effect of choices on the spatial structure of the product is useful (Kronenfeld, 2003; Thompson, 2005).

Third, network theory itself suggests a number of potentially fruitful research themes, including the analysis of: fractal dimensions of the patches and the networks (Barabasi and Albert, 1999), connectivity and characterization of the larger components of LC classes, distance properties (Newman *et al.*, 2001), fragmentation and cut points (Urban and Keitt, 2001), spatial autocorrelation (Atkinson and Tate, 2000), and chromatic properties (Chartrand and Zhang, 2005; Kearns *et al.*, 2006). Those themes that bear most upon salient environmental issues should be given highest priority.

Fourth, perhaps the most important problem in LC analysis is the characterization and forecasting of change (Comber *et al.*, 2004) in order to understand interactions among landscape states and other phenomena, for example urbanization and water quality (Claggett *et al.*, 1996; Stehman et al., 2003). A key issue in change analysis is the distinction between reliability and accuracy (Loveland *et al.*, 2002) in order to reliably partition temporal variation among such factors as the nature of the data gathering, processing algorithms, technology, and parameterization techniques versus actual changes on the landscape. The all-important distance matrix describing degrees of separation represent interactions among classes that can tell us much about process (Vogt *et al.*, 2007).

Finally, an obvious next step is to examine the dynamic behavior of networks, and there are at least three lines of attack. We can compare the network descriptions across time as is done in Figure 10. We can create new networks out of raster overlays representing patches where pixels have or have not made transitions among LC types. And, we can create space/time networks in which links represent both spatial adjacencies (what are the LC codes of neighboring nodes?(Cressie, 1991)) as well as temporal persistence (what are the codes of nodes joined by time steps?), so that patches now span time as well as space (Quattrochi and Goodchild, 1996) (Grimm *et al.*, 2005). Judicious sampling makes this a straightforward extension of what has already been done.

Challenges

Although the land-cover network system has evident benefits, it faces at least four challenges. First, image-based analysis must always focus on a more or less arbitrary window (in this case $1{,}024^2$ cells) that risks truncating the largest phenomena. Future studies will apply these techniques to windows at various locations and of different sizes, not to mention different resolutions.

A second challenge is the sampling system, which requires sensitivity analysis. In practical raster analysis a given resolution may be appropriate, but "noise" will often be filtered, say by assigning to patches below a certain size the value of a surrounding patch. The present report is based on a censored stratified sample of the largest patches, so more attention must be given to the implications of this kind of filtering (Gallant *et al.*, 2004). Any attempt to reduce noise inevitably degrades signal, but it is unlikely that patch analysts will retain *all* of the patches. Preliminary results suggest that within wide limits the key network descriptions (statistics, distances, autocorrelation) are robust, i.e., not greatly sensitive to judicious sampling; but more fragmented LC classes, which often present detection problems or critical policy issues, will inevitably be undersampled. Because combinatorial operations are inherently computationally intensive, sampling may be desirable for larger rasters, but the inexorable growth of computational power lessens this motivation.

Finally, on an ontological note, this research raises persistent questions of what phenomenon LC data describe (Mark *et al.*, 1999), what we are analyzing (Comber *et al.*, 2004), and how LC models predict other phenomena. The present technique offers the possibility of bridging the vexing gap between the "stuff" in an image and the "things" in a spatial database; note for instance how some of the largest nodes in Plate 1 were readily named.

In spite of these challenges, this research provides opportunities of linking LC data to the ecological processes driving and being driven by landscape status and trends (Ricotta *et al.*, 2000), suggesting the following directions. Given that the streams and watersheds of hydrologic systems are themselves tree networks, it is useful to associate water quality data with LC networks (Amoros and Bornette, 2002). These planar networks can also be conflated with network representations of animal habitats (Urban *et al.*, 2006), migration paths *(Farnsworth et al.*, 2004; Alerstam, 2006), contagion webs (Viboud *et al.*, 2006) and many other highly linked biogeophysical processes (Dietrich and Taylor, 2006).

Knowledge of landscape dynamics is fundamental to understanding Earth systems dynamics, which is the fundamental problem of the twenty-first century (National Research Council, 2001). This knowledge is hard to acquire because natural phenomena manifest highly parallel relationships across time and space at all scales (Wolfram, 2002). Yet the data of geography are becoming rich enough (and geographic methods powerful enough) to model this complexity using simple structures that capture some of the infinite richness of a networked world.

References

Alerstam, T., 2006. Conflicting evidence about long-distance animal navigation, *Science*, 313:791–794.

Amoros, C., and G. Bornette, 2002. Connectivity and biocomplexity in waterbodies of riverine floodplains, *Freshwater Biology*, 47(4):761–776.

Atkinson, P.M., and N.J. Tate, 2000. Spatial scale problems and geostatistical solutions: A review, *Professional Geographer*, 52(4):607–623.

Barabasi, A.-L., and R. Albert, 1999. Emergence of scaling in random networks, *Science*, 286:509–512.

Batty, M., 2005. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-based Models, and Fractals*, MIT Press, Cambridge, Mass., xxi, 565 p.

Bettencourt, L.M.A., J. Lobo, D. Helbing, C. Kuhnert, and G.B. West, 2007. Growth, innovation, scaling, and the pace of life in cities, *Proceeding of the National Academy of Sciences*, 104(17):7301–7306.

Boots, B., 2003. Developing local measures of spatial association for categorical data, *Journal of Geographical Systems*, 5(2):139–160.

Bunn, A.G., D.L. Urban, and T.H. Keitt, 2000. Landscape connectivity: A conservation application of graph theory, *Journal of Environmental Management*, 59:265–278.

Chartrand, G., and P. Zhang, 2005. *Introduction to Graph Theory*, McGraw-Hill Higher Education, Boston, xii, 449 p.

Claggett, P.R., C.A. Jantz, S.J. Goetz, and C. Bisland, 1996. Assessing development pressure in the Chesapeake Bay watershed: An evaluation of two land-use change models, *Environmental Monitoring and Assessment*, 94(1-3):129–146.

Comber, A., P. Fisher, and R. Wadsworth, 2004. Integrating land-cover data with different ontologies: Identifying change from inconsistency, *International Journal of Geographical Information Science*, 18(7):691–708.

Cova, T.J., and M.F. Goodchild, 2002. Extending geographical representation to include fields of spatial objects, *International Journal of Geographical Information Science*, 16(6):509–532.

Cressie, N.A.C., 1991. *Statistics for Spatial Data*, Wiley, New York, 900 p.

Cushman, S.A., K.e. McGarigal, and M.C. Neel, 2008. Parsimony in landscape metrics: Strength, universality, and consistency, *Ecological Indicators*, 8:691–703.

De'ath, G., and K.E. Fabricius, 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis, *Ecology*, 81(11):3178–3192.

De Cola, L., 1989. Fractal analysis of a classified Landsat scene, *Photogrammetric Engineering & Remote Sensing*, 55(5):601–610.

de Smith, M.J., M.F. Goodchild, and P. Longley, 2007 *Geospatial Analysis*, Troubadore, Leicester UK, 394 p.

Dietrich, W.E., and P.J. Taylor, 2006. The search for a topographic signature of life, *Nature*, 439:411–418.

Environmental Protection Agency, 2002. *Mid-Atlantic Land-cover Change*, EPA Report EPA/600/R-02/035.

Environmental Systems Research Institute, 2006. ArcGIS, Geographic Information System.

Farnsworth, A., S.A. Gauthreaux, and D. von Blaricom, 2004. A comparison of nocturnal call counts of migrating birds and reflectivity measurements on Doppler radar, *Journal of Avian Biology*, 35:365–369.

Ferraz, G. J.D. Nichols, J.E. Hines, P.C. Stouffer, R.O. Bierregaard, Jr., and T.E. Lovejoy, 2007. A large-scale deforestation experiment: Effects of patch area and isolation on Amazon birds, *Science*, 315:238–241.

Fortin, M.-J., and M. Dale, 2005. *Spatial Analysis: A Guide for Ecologists*, Cambridge University Press, Cambridge UK.

Gallant, A.L., T.R. Loveland, T.L. Sohl, and D.E. Napton, 2004. Using an ecoregion framework to analyze land-cover and land-use dynamics, *Environmental Management*, 34(Supplement 1):S89–S110.

Gentleman, R., 2005. *Bioinformatics and Computational Biology Solutions using R and Bioconductor,* Springer Science+Business Media, New York, xix, 473 p.

Griffith, J.A., S.V. Stehman, and T.R. Loveland, 2003. Landscape trends in the Mid-Atlantic and Southeastern United States ecoregions, *Environmental Management*, 32(5):572–588.

Grimm, V., and D.L. DeAngelis, 2005. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology, *Science*, 310:987–991.

Guptill, S.C., 1990. *An Enhanced Digital Line Graph Design: A Feature-based Data Model for Digital Spatial Data Bases that Represents Geographic Phenomenon*, U.S. Geological Survey, Circular 1048, Reston Virginia 148 p.

Haining, R.P., 2003. *Spatial Data Analysis : Theory and Practice*, Cambridge University Press, Cambridge, UK, New York, xx, 432 p.

Helsel, D.R., and R.M. Hirsch, 1992. *Statistical Methods in Water Resources*, Elsevier, Amsterdam, New York, xvi, 522 p.

Hilhorst, H.J., 2005. Planar Voronoi cells and the failure of Aboav's law, *arXiv*, 1(0509409).

Hinton, G.E., and R.R. Salakhutdinov, 2006. Reducing the dimensionality of data with neural networks, *Science*, 313:504–507.

Hyndman, R.J., A.B. Koehler, R.D. Snyder, and S. Grose, 2002. A state space framework for automatic forecasting using exponential smoothing methods, *International Journal of Forecasting*, 18(3):439–454.

Insightful Corporation, 2001 *S-Plus 6 for Windows Programmer's Guide*, Seattle, Washington, 988 p.

Jones, K.B., C.M. Edmonds, E.T. Slonecker, J.D. Wickham, A.C. Neale, T.G. Wade, K.H. Riitters, and W.G. Kepner, 2008. Detecting changes in riparian habitat conditions based on patterns of greenness change, *Ecological Indicators*, 8:89–99.

Jones, K.B., A.C. Neale, T.G. Wade, J.D. Wickham, C.L. Cross, C.M. Edmonds, T.R. Loveland, M.S. Nash, K.H. Riitters, and E.R. Smith, 2001. The consequences of landscape change on ecological resources: An assessment of the United States Mid-Atlantic Region, 1973–1993, *Ecosystem Health*, 7: 229–242.

Kearns, M., S. Suri, and N. Monfort, 2006. An experimental study of the coloring problem on human subject networks, *Science*, 313:824–827.

Kronenfeld, B., 2003. Implications of a data reduction framework to assignment of fuzzy membership values in continuous class maps, *Spatial Cognition & Computation*, 3(n/3): 221–238.

Loveland, T.R., T.L. Sohl, S.V. Stehman, A.L. Gallant, K.L. Sayler, and D.E. Napton, 2002. A strategy for estimating the rates of recent United States land-cover changes, *Photogrammetric Engineering & Remote Sensing*, 68(10):1091–1099.

Mark, D.M., B. Smith, and B. Tversky, 1999. Ontology and geographic objects: An empirical study of cognitive categorization, *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science, Proceedings of the International Conference COSIT'99,* Stade, Germany, August, 1661/1999:283–298.

National Research Council (US) Committee on Grand Challenges in Environmental Sciences, 2001. *Grand Challenges in Environmental Sciences*, National Academy Press, Washington, D.C., 96 p.

Newman, M.E.J., S.H. Strogatz, and D.J. Watts, 2001. Random graphs with arbitrary degree distributions and their applications, *Physical Review E*, E64(2).

Pascual-Hortal, L., and S. Saura, 2006. Comparison and development of new graph-based landscape connectivity indices: towards the priorization of habitat patches and corridors for conservation, *Landscape Ecology*, 21(7):959–967.

Pemmaraju, S., and S. Skiena, 2003. *Computational Discrete Mathematics: Combinatorics and Graph Theory with Mathematica*, Cambridge University Press, Cambridge, England, 480 p.

Quattrochi, D.A., and M.F. Goodchild, 1996. *Scale in Remote Sensing and GIS*, Lewis, Boca Raton, Florida, 406 p.

Richards, J.A., and X. Jia, 2006. *Remote Sensing Digital Image Analysis: An Introduction*, Springer, Berlin, 439 p.

Ricotta, C., A. Stanisci, G.C. Avena, and C. Blasi, 2000. Quantifying the network connectivity of landscape mosaics: A graph-theoretical approach, *Community Ecology*, 1(1):89–94.

Riitters, K.H., J.D. Wickham, R.V. O'Neill, K.B. Jones, E.R. Smith, J.W. Coulston, T.G. Wade, and J.H. Smith, 2002. Fragmentation of continental United States forests, *Ecosystems*, 5(8):815–822.

Riitters, K.H., J.D. Wickham, and T.G. Wade, 2006. Evaluating ecoregions for sampling and mapping land-cover patterns, *Photogrammetric Engineering & Remote Sensing*, 72(7):781–788.

Samet, H., 1990. *The Design and Analysis of Spatial Data Structures,* Addison-Wesley, Reading Massachusetts, 493 p.

Sohl, T.L., K.L. Sayler, M.A. Drummond, and T.R. Loveland, 2007. The FORE-SCE model: A practical approach for projecting land-cover change using scenario-based modeling, *Journal of Land Use Science*, 2(2):10–126.

Stauffer, D., and A. Aharony, 1992. *Introduction to Percolation Theory*, Taylor & Francis, London, Washington, D.C., 181 p.

Stehman, S.V., T.L. Sohl, and T.R. Loveland, 2003. Statistical sampling to characterize recent United States land-cover change, *Remote Sensing of Environment*, 86:517–529.

Thompson, J.N., 2005. *The Geographic Mosaic of Coevolution*, University of Chicago Press, Chicago, 443 p.

Turner, M.J., J.M. Blackledge, and P.R. Andrews, 1998. *Fractal Geometry in Digital Imaging*, Academic Press, San Diego, California, 328 p.

Urban, D.L., and T.H. Keitt, 2001. Landscape connectivity: A graph-theoretic perspective, *Ecology*, 82(5):1205–1218.

Urban, D.L., R.I. McDonald, E.S. Minor, and E.A. Treml, 2006. *Book Scaling and Uncertainty Analysis in Ecology* (J. Wu, K.B. Jones, H. Li and O.L. Loucks, editors), Dordrecht, The Netherlands, pp. 239–257.

Venables, W.N., and B.D. Ripley, 1999. *Modern Applied Statistics with S-PLUS*, Springer-Verlag, New York, 501 p.

Viboud, C., N.B. Ottar, L.S. David, S. Lone, A.M. Mark, and T.G. Bryan, 2006. Synchrony, waves, and spatial hierarchies in the spread of influenza, *Science*, 312(5772):447–451.

Vogt, P., K.H. Riitters, C. Estregull, J. Kozak, T.G. Wade, and J.D. Wickham, 2007. Mapping spatial patterns with morphological image processing, *Landscape Ecology*, 22:171–177.

Wickham, J.D., K. Riitters, T.G. Wade, M. Coan, and C. Homer, 2007. The effect of Appalachian mountaintop mining on interior forest, *Landscape Ecology*, 22:179–187.

Wickham, J.D., K.H. Riitters, R.V. O'Neill, K.B. Jones, and T.G. Wade, 1996. Landscape 'contagion' in raster and vector environments, *International Journal of Geographical Information Systems*, 10(7):891–899.

Wolfram, S., 2002. *A New Kind of Science*, Wolfram Media, Champaign Illinois, 1197 p.

Worboys, M., and M. Duckham, 2004. *GIS: A Computing Perspective*, CRC Press, Boca Raton, Florida, 426 p.

Yang, L., G. Xian, J.M. Klaver, and B. Deal, 2003. Urban land-cover change detection through sub-pixel imperviousness mapping using remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 69(9):1003–1010.

(Received 18 August 2008; accepted 11 December 2008; final version 16 June 2009)